How persuasive is phishing email? The role of authentic design, influence and current events in email judgements

Emma J Williams

School of Management, University of Bath, UK

Danielle Polage

School of Psychology, Central Washington University

Abstract

Fraudulent emails, otherwise known as phishing emails, use a range of influence techniques to persuade individuals to respond, such as promising a monetary reward or invoking a sense of urgency. The current study explored a number of factors that may impact the persuasiveness and trustworthiness of emails by examining participant judgements of 20 pre-designed emails that varied according to (a) whether they used *loss or reward-based influence* techniques, (b) whether they contained particular *authentic design* cues, (c) whether they referenced a *salient* current event (the Rio Olympics), and (d) whether participants had been previously exposed to information regarding online scams in general. Results suggest that the presence of authentic design cues and the type of influence technique used significantly impacted participant judgements. Findings are discussed in relation to theoretical models of phishing susceptibility.

Keywords: influence, phishing, online trust, persuasive communications, cyber security

1. Introduction

Phishing emails are a growing menace across society, providing a quick and easy way to target large numbers of users with fraudulent messages. Within organisations, approximately one in ten employees have been found to click on links or open attachments during internal phishing tests (Verizon, 2016). The use of familiar logos and branding, accurate email layouts, and spoofing of email addresses to appear authentic, means that it is increasingly difficult to differentiate phishing emails from legitimate emails in the modern age. But what is it that makes people suspicious of emails that they receive? And what is it that persuades them to respond?

Current knowledge and understanding regarding how people make decisions regarding the legitimacy and persuasiveness of emails that they receive remains limited. In particular, how phishing emails use particular content within a message to influence the susceptibility of recipients. Addressing this limitation is vital if the evolving problem of phishing scams is to be addressed. Investigating this issue is the primary aim of the current study. First, we provide an overview of previous literature related to how people evaluate emails that they receive, before outlining the aims and hypotheses of the current study. We then present our methodological approach and results. Finally, we consider our findings in relation to current theories of susceptibility to phishing emails.

1.1. Theoretical Background

1.1.1. Role of Message-specific Factors

Previous work examining the factors that impact susceptibility to phishing emails has ranged from sending participants pre-designed simulated phishing emails (e.g., Wright & Marett, 2010) to examining participant judgements of screenshots of such emails (e.g., Canfield, Fischhoff & Davis, 2016; Pattinson, Jerram, Parsons,

McCormac, & Butavicius, 2012). The majority of these studies have focused on exploring the role of demographic or personality related factors, or the impact of increased knowledge and training interventions, on susceptibility, rather than the relative impact of various message-related factors per se.

Within the integrated information processing model of phishing susceptibility, however, Vishwanath, Herath, Chen, Wang and Rao (2011) consider the use of particular influence techniques on susceptibility to phishing. Influence techniques are techniques used within the email content to persuade people to respond. They include instilling a sense of urgency by invoking deadlines or time-limited offers, offering some form of reward, or suggesting that failure to respond to the email will incur a loss to the individual (Atkins & Huang, 2013; Cialdini, 2007; Stajano & Wilson, 2011; Workman, 2008). Vishwanath et al. (2011) claimed that the presence of influence techniques within a phishing email monopolises people's limited attentional resources, leading to other elements within the message, such as an incorrect sender address, being overlooked. This is considered to occur because influence techniques encourage more automatic forms of information processing that are reliant on the use of mental shortcuts (known as *heuristics*). As a result, more systematic consideration of the legitimacy of a potential message is not undertaken. The use of such heuristic processing strategies when evaluating information has been linked with a greater likelihood of considering information to be genuine (termed *truth-bias*; Harrison, Vishwanath & Rao, 2016; Levine, 2014; Toma & Hancock, 2012; Williams, Morgan & Joinson, 2017).

Whether particular influence techniques are more effective than others in encouraging these responses, however, is unclear. Although previous research has tentatively suggested that people are differentially susceptible to certain influence

techniques (e.g., Butavicius, Parsons, Pattinson & McCormac, 2015; Oliveira, Rocha, Yang, Ellis, et al., 2017), current understanding in this area remains limited. For instance, phishing messages commonly either offer individuals a reward to encourage responses or suggest a loss will be incurred if there is a failure to respond (such as freezing access to an online account), but little examination of the relative impact of these particular techniques has been undertaken. Although Harrison, Svetieva and Vishwanath (2016) recently failed to show a difference in susceptibility to emails that contained threat versus reward-based cues, the psychological concept of loss aversion suggests that individuals are likely to be more sensitive to potential losses than gains (Kahneman & Tversky, 1984). If this is the case, then emails that reference a potential loss may be more persuasive than those that focus on potential rewards. Although these influence techniques may not directly influence perceived trustworthiness of an email, however, the pressure to maintain consistency in responses and avoid cognitive dissonance in thoughts and beliefs (Festinger, 1962), may result in emails that are considered more persuasive also being rated as more trustworthy. Further work is required to examine this possibility.

When making decisions regarding the legitimacy of online information more generally, a range of factors has been suggested to influence people's judgements of the relative trustworthiness of a communication. Similar to the influence literature discussed above, these mainly occur as a result of people's limited ability to process information. For example, individuals have been found to use relatively superficial cues related to the 'look and feel' of information when determining trustworthiness (Sillence, Briggs, Harris, & Fishwick, 2006; Fogg, Marshall, Ospiovich, Varma, et al., 2000). Similarly, people are likely to only process the most noticeable (or *salient*) features within a message when evaluating a communication (Lang, 2000). Finally,

heuristics and biases related to the perceived credibility of a message source, and whether information violates expectations, also influence judgements (Metzger, Flanagin & Medders, 2010). Unfortunately, as phishing emails become more sophisticated in their design, they are increasingly able to mimic established and reputable brands and use expected email layouts in order to appear more credible. The extent to which people use these design aspects when determining the trustworthiness of email messages, however, remains uncertain. Furthermore, current understanding of how the presence of particular influence techniques within the message content may impact these judgements is also unclear.

In addition to influence techniques and design elements, recent examples of phishing attacks have emerged whereby online scams reference well-known current events in an effort to appear more legitimate. For instance, following the high-profile data breach suffered by the TalkTalk Telecom Group in the UK, whereby hackers accessed sensitive customer information, a number of phishing scams emerged related to this event, with customers being targeted by scams that referenced the recent data breach. Similarly, scams linked to current sporting events and religious festivals have also been highlighted (BBC News, 2017). Linking phishing emails to events that individuals are likely to already be familiar with may further enhance the perceived credibility of the email, with previously encountered information likely to be more salient within memory and therefore more likely to be considered genuine due to this ease of accessibility (Begg, Anas & Farinacci, 1992). In this way, exposure to current events via media reports may make these events more salient, effectively priming recipients (see Meyer & Schvaneveldt, 1971) for later phishing emails that reference these events. However, whether linking phishing emails to salient events influences people's perceptions of the trustworthiness and persuasiveness of emails has not yet

been explored. Understanding how these various message-related factors (i.e., influence techniques, design cues, and referencing salient current events) impact perceptions of emails is the primary aim of the current study.

1.1.2. Role of External Factors

Previous considerations of phishing susceptibility have also suggested that people who are more aware of the potential risks of online environments may be less susceptible to such emails (Vishwanath, Harrison & Ng, 2016). Although the use of educational and game-based interventions has previously been reported to increase people's awareness of the risks of phishing emails (e.g., Arachchilage & Love, 2014; Kumaraguru, Rhee, Sheng, Hasan, et al., 2007; Sun & Chen, 2016), the impact of more general information in the wider environment on perceptions of phishing emails has not yet been examined. It is important that this is addressed, since members of the public are increasingly exposed to media stories within the local or national news regarding current online threats and the victims of online scams (e.g., BBC News, 2016). Whether such media reports influence perceptions of email communications in any way, however, is currently unknown. A secondary aim of this paper, therefore, is to examine the impact of immediate prior exposure to general media reports regarding the risk of online scams on subsequent email judgements.

Finally, people who have developed habitual responses to email communications (e.g., they engage with emails automatically and are thus less able to control their behaviour) have been suggested to be more susceptible to phishing attacks (Vishwanath, 2015). However, it is currently unknown whether habitual email behaviour also impacts how people evaluate email communications. For instance, are habitual email responders also more likely to assume that emails are trustworthy? The final aim of this paper is to address this question.

1.2. Hypotheses Development

This study examines the impact of particular message-related factors (specifically, influence techniques, authentic design cues, and reference to salient current events) on individual judgements of the perceived trustworthiness and persuasiveness of emails. These message factors are predominantly based on aspects that have previously been considered likely to influence people's susceptibility to phishing emails. However, the relationship between these aspects, and how they may differentially influence judgements of emails, is not currently clear. By systematically manipulating and assessing the relative contribution of these various message-related factors, it will be possible to address this current limitation.

Hypothesis One: The concept of loss aversion suggests that people are more sensitive to losses than gains (Kahneman & Tversky, 1984). If loss-based techniques are more persuasive than reward-based ones, loss-based emails will be rated as more persuasive (H1a) than reward-based emails. Similarly, considering loss-based emails more persuasive may also result in greater ratings of trustworthiness (H1b) and a greater likelihood of choosing to respond to such emails (H1c).

Hypothesis Two: Previous research related to judgements of online information suggests that design cues, such as the presence of a logo, influence the perceived legitimacy of information (e.g., Sillence et al., 2006). We propose that emails that contain particular design cues, such as the presence of a logo and a copyright statement, will be rated as more trustworthy (H2a) and more persuasive (H2b) than those that do not. These emails will also be rated as more likely to be responded to (H2c).

Hypothesis Three: Previous research suggests that previously encountered information is more likely to be considered genuine (Begg et al., 1992). Phishing

emails have also been known to reference current events, however the impact of referencing salient event information on judgements of emails has yet to be examined. We propose that emails that reference a salient current event will be rated as more trustworthy (H3a) and more persuasive (H3b) than emails that do not. We also propose that participants will rate these emails as more likely to be responded to (H3c).

In addition to these message-related factors, our study also investigates the extent to which email judgements may be influenced by prior exposure to media information regarding the risks of online scams. The impact of generic media stories has yet to be examined, despite media agencies increasingly reporting stories related to these risks. It is important, therefore, to consider the role that such information may have in influencing how people evaluate various types of emails.

Hypothesis Four: We propose that participants who are exposed to media stories regarding the risks of online scams will rate emails as less trustworthy (H4) compared to participants who are not exposed to this information, due to temporarily increased levels of generalised suspicion.

Previous research has also suggested that habitual online behaviour (i.e., automatically responding to emails received or clicking on links) increases susceptibility to phishing (Vishwanath, 2015). However, it is currently uncertain whether such automatic responses are also related to greater trust in email communications in general.

Hypothesis Five: If habitual email behaviour is also related to an increased likelihood of considering information to be trustworthy, then a positive correlation will be found between self-reported habitual email use and ratings of trust (H5).

Finally, within this study we also explore whether participant perceptions of their decision making strategies correspond with the factors listed above (i.e., the presence of influence techniques and design cues, and the role of routine behaviour).

By addressing these research questions, we hope to contribute to the further development of theoretical models in this area. Specifically, we aim to further understanding in relation to the role of loss and reward-based influence techniques, design cues, and referencing salient current events, on evaluations of email trustworthiness and persuasiveness. This work also contributes to current understanding regarding the potential role of particular external factors on email judgements, particularly the extent to which generic awareness messages influence email judgements. By understanding what information does, and does not, impact perceptions of phishing emails, it will be possible to develop better awareness communications in the future.

2. Method

A 2x2x2x2 mixed design was used, with three within group factors and one between group factor. The within group factors included (a) whether the email included loss or reward-based influence techniques, (b) whether the email included specific authentic design cues (i.e., the email contained particular design cues vs. did not contain these cues), and (c) whether the email referenced a salient current event (i.e., the email referenced the Rio Olympics vs. made no reference to the Rio Olympics). The between group factor was whether participants were exposed to the online risks news story or to an unrelated news story prior to the email judgement task. Dependent variables included how trustworthy and persuasive the emails were considered to be and whether participants would respond to the email.

2.1. Stimuli and Measures

2.1.1. Risk Information Manipulation

A brief news story regarding the risk of online scams (total 80 words) and an unrelated news story that did not reference online scams (total 82 words) were created. The risk of online scams story represented our generic risk awareness manipulation. News stories were designed by the researchers and based on actual current news stories. Stories were matched according to layout, style and length and are shown in the Appendix.

2.1.2. Salient Current Event

A brief news story regarding the closing ceremony of the 2016 Rio Olympics (total 74 words) was also created based on actual current news stories. This was to ensure that all participants were familiar with this particular current event, which was then used as the salient current event referenced within a subset of emails. This story is shown in the Appendix.

2.1.3. Email Stimuli

The email stimuli used within the study consisted of twenty pre-designed emails that varied according to the presence of particular cues. Although the researchers designed these emails, the content and layout was based on phishing emails received by the authors or found online. This is an approach previously used in phishing susceptibility research to identify email stimuli (e.g., Parsons, McCormac, Pattinson, Butavicius & Jerram, 2013). This approach enabled us to systematically manipulate the presence of particular cues within the emails and therefore investigate how these cues influence judgements of emails.

Emails were designed to vary on the following dimensions:

1. **Influence technique:** Emails varied according to whether they focused on reward-based or loss-based influence techniques. Reward-based emails

offered some form of reward for responding, such as a free gift or special offer (e.g., ‘Just click HERE to be in with a chance to win!’). Loss-based emails suggested that the recipient would lose access to something if they failed to respond, such as an account being frozen (e.g., ‘Your internet banking access has been temporarily suspended… Click HERE to reinstate your account’). Loss-based emails did not reference the Rio Olympics.

2. **Authentic design cues:** Emails varied according to whether they contained authentic design cues or not. Those containing authentic design cues included a relevant logo, a purported copyright statement at the bottom of the email (e.g., ‘Copyright © 2016 [company name]. All Rights Reserved.’) and reference to the professed organisation within the purported sender email address (e.g., From: [company name] (do_not_reply@euro.companyname.com) rather than a generic email address (e.g., ‘freeiphone@gmail.com’). All email addresses were inaccurate, however, and should therefore have invoked a degree of suspicion in participants. Emails were identical in every other way.
3. **Salient current event reference:** Emails varied according to whether they referenced a salient current event (specifically, the 2016 Rio Olympics) within the message content (e.g., ‘[company name] is actively supporting the Rio Olympics this summer. To celebrate this we are offering…’) and email subject heading (e.g., ‘Win free flights to celebrate Rio 2016!’) or did not reference any particular event. Emails that referenced Rio 2016 and those that did not were identical in all other aspects to enable matched comparison. The same professed organisations were also used within both sets of emails in order to prevent familiarity with a particular organisation confounding the results.

To enable us to examine the influence of these various factors without requiring participants to view too many emails, only four loss-based emails were included. These did not reference the Rio Olympics and were used as a comparator with the first four reward-based emails viewed by participants that did not reference the Rio Olympics. All emails were matched according to length, layout, and grammar / spelling, with emails only differing according to the particular aspects described above. All emails referenced well-known organisations that have been used in phishing attacks. In sum, this provided the following six conditions. The first four conditions all used reward-based influence techniques:

1. Four high authentic design emails that referenced the Rio Olympics (Mean word count = 69.75; *SD* =2.87);
2. Four high authentic design emails that did not reference the Rio Olympics (Mean word count = 69.00; *SD* =3.83);
3. Four low authentic design emails that referenced the Rio Olympics (Mean word count = 62.25; *SD* =3.30);
4. Four low authentic design emails that did not reference the Rio Olympics (Mean word count = 61.75; *SD* =3.40);
5. Two high authentic design emails that used loss-based influence techniques (Mean word count = 60.00; *SD* =2.83);
6. Two low authentic design emails that used loss-based influence techniques (Mean word count = 53.00; *SD* =1.41).

Example stimuli for each condition are shown in the Appendix. As we were interested in understanding what message factors make people more or less suspicious of various cues commonly contained within phishing emails, rather than their ability to

discriminate between phishing and legitimate emails per se, we did not include any legitimate emails for comparison.

2.1.4. Questionnaire Measures

The 12-item email habits questionnaire (Vishwanath et al., 2016) was used to provide a measure of habitual email use. This self-report questionnaire measure is an adapted version of the Self-Report Habit Index (Verplanken & Orbell, 2003) and measures the extent that people engage in habitual, automatic behaviour in a particular context. In this case, whether individuals engage in habitual use of email. For example: 'Email use is something I do without thinking' (answered on a scale of 1-7).

A number of additional questions were also included within the study. These included a rating of familiarity for each of the organisations referenced within the phishing emails (on a scale of 1-5; 1 = extremely familiar, 5 = not at all familiar) and three questions taken from Williams et al. (2017) related to general awareness of scams: 'To what extent do you trust email communications in general?' (1 = almost never, 7 = almost always); 'How confident are you in your ability to differentiate genuine communications from scam communications in daily life?' (1 = very uncertain, 7 = very confident); and 'How would you rate your awareness of the common techniques used in scams?' (1 = very unaware, 7 = very aware).

2.2. Data Collection

2.2.1. Participants

Participants were recruited from Central Washington University to complete an online study exploring mass-marketing communications. One hundred and eighty-four participants completed the study in exchange for university course credit. Six participants were excluded due to a failure to complete the judgement task, resulting in a final sample size of 178. One hundred and twenty-four participants were female

and fifty-four were male. The majority of participants reported being in the 18-24 years age group category (168 participants), the remaining 10 participants were over the age of 24.

2.2.2. Procedure

The University's Ethical Review Board granted ethical approval for the study. The study was conducted online using the Qualtrics online survey platform (www.qualtrics.com). Informed consent was provided online prior to the main study, with participants provided with ethical information regarding study participation before being required to complete a check box in order to access the main survey. Following this, participants provided demographic information (e.g., age, gender, nationality, employment status and education level) before being randomly allocated to either the risk information group (90 participants) or the control group (88 participants) by the survey software. *Exposure to risk information: between-group manipulation:* Prior to the email judgement task, those in the risk information group were exposed to the risk of online scams news story. Those in the control group viewed the unrelated news story.. To ensure that participants attended to the news stories, they were asked 'What is the subject of this news story?' following exposure to the story (open-ended question). All participants completed this question accurately.

Exposure to salient current event: Following this, all participants were exposed to the 2016 Rio Olympics news story. Once again, to ensure that participants attended to the news story, they were asked 'What is the subject of this news story?'. All participants completed this question accurately.

Email judgement task: Participants then viewed the series of 20 pre-designed emails. For each email, participants were required to indicate (a) whether they would be most

likely to respond to the email or ignore the email (Respond (R)/ Ignore (I); recorded as 1 for respond and 2 for ignore within the dataset), (b) how trustworthy they considered the email to be (on a scale of 1-5, where 1 = extremely trustworthy and 5 = not at all trustworthy) and (c) how persuasive they considered the email to be (on a scale of 1-5, where 1 = extremely persuasive and 5 = not at all persuasive). Following the email judgement task, participants completed the questionnaire measures. Finally, participants were asked what factors influenced their decision regarding whether to respond to an email. This was an open-ended question that provided qualitative data for thematic analysis. Following this, participants were automatically directed to full debriefing information regarding the study. Data collection took place in Autumn 2016.

2.3. Data Analysis

2.3.1. Quantitative Data: Email Judgements

To examine the influence of message-related factors and risk information on email judgements, mixed ANOVAs were computed for each of the dependent variables (persuasiveness ratings, trust ratings, and response choice (the proportion of emails to which participants chose 'respond' for each condition)). Primary influence technique (reward vs. loss), authentic design cues (high vs. low) and reference to salient current events (reference Rio Olympics vs. no reference to Rio Olympics) were used as within group factors and group (risk information group vs. control group) was used as the between group factor.

To examine whether habitual email use was related to email judgements, bivariate correlations were computed between scores on the adapted Self-Report Habit Index (Verplanken & Orbell, 2003; Vishwanath et al., 2016) and mean trust

ratings, persuasiveness ratings and response choice across the various email conditions.

2.3.2. Qualitative Data: Participants Self-reported Reasons for their Decisions

A single rater analysed participants' self-reported reasons for their decisions according to the presence of particular theoretically-driven themes based on the study hypotheses and previous research regarding susceptibility to phishing (Stajano, & Wilson, 2011; Vishwanath et al., 2011; 2016; Williams et al., 2017). These themes included:

(a) *The role of influence techniques* - reference to the presence of particular influence techniques included within emails as influencing their decision, such as known and trusted organisations (authority influence technique), the requirement to undertake an urgent or important action quickly (urgency influence technique), the provision of incentives and rewards (reward-based influence technique), the threat of loss (loss-based influence technique), or referencing a salient current event.

(b) *The presence of authentic design cues* - reference to particular authentic design cues present within the email, such as the use of a logo, the use of copyright statements, and email addresses that appear legitimate.

(c) *Individual habits and routine behaviour* - reference to behavioural habits, such as always ignoring or responding to particular types of emails.

3. Results

First, we present the findings related to the message-specific factors (H1-H3) before presenting results related to the between group factor of exposure to information regarding online risks (H4). We then present the results related to the email habits questionnaire (H5) and results of the qualitative data analysis. Table 1 provides

summary statistics according to each condition and Table 2 summarises results in relation to study hypotheses.

Table 1. Summary statistics of ratings of email trust, persuasiveness and response likelihood according to group and condition.

		Risk Information Group			Control Group		
		Trust ratings	Persuasiveness ratings	Response likelihood	Trust ratings	Persuasiveness ratings	Response likelihood
High Authentic Design Cue	Reward: Reference Rio Olympics	3.61(1.00)	3.51(.93)	23%	3.38(1.08)	3.37(1.00)	33%
	Reward: Does not reference Rio Olympics	3.55(.81)	3.51(.82)	28%	3.49(.84)	3.45(.76)	33%
	Loss	3.23(1.08)	3.08(1.05)	47%	3.16(.97)	3.02(.99)	57%
Low Authentic Design Cue	Reward: References Rio Olympics	4.58(.54)	4.37(.68)	4%	4.47(.63)	4.23(.71)	8%
	Reward: Does not reference Rio Olympics	4.49(.62)	4.27(.71)	7%	4.38(.64)	4.23(.69)	9%
	Loss	4.40(.84)	4.08(.95)	19%	4.33(.90)	3.93(1.11)	24%

Note: Due to the likert scale anchors used, lower trust and persuasiveness ratings = increased trust and persuasiveness.

Table 2. Summary of findings in relation to study hypotheses.

H1a: Loss-based emails will be rated as more persuasive than reward-based emails.	Supported
H1b: Loss-based emails will be rated as more trustworthy than reward-based emails.	Supported
H1c: Loss-based emails will be more likely to be rated respond than reward-based emails.	Supported
H2a: Emails that contain particular design cues, such as the presence of a logo and copyright statement, will be rated as more trustworthy than those that do not.	Supported
H2b: Emails that contain particular design cues, such as the presence of a logo and copyright statement, will be rated as more persuasive than those that do not.	Supported
H2c: Emails that contain particular design cues, such as the presence of a logo and copyright statement, will be more likely to be rated respond than those that do not.	Supported
H3a: Emails that reference a salient current event will be rated as more trustworthy than emails that do not.	Not supported
H3b: Emails that reference a salient current event will be rated as more persuasive than emails that do not.	Not supported
H3c: Emails that reference a salient current event will be more likely to be rated respond than emails that do not.	Not supported
H4: Participants who are exposed to generic information regarding online risks will rate emails as less trustworthy compared to participants not exposed to this information.	Not supported
H5: Self-reported habitual email use will be related to an increased likelihood of considering emails to be trustworthy.	Not supported

3.1. H1: The Impact of Loss and Reward-based Influence Techniques on Email Judgements

H1a: In relation to email persuasiveness, a main effect of influence technique was found, $F(1, 176) = 31.46$, $p = .001$, $\eta 2 = .05$ such that emails using loss-based influence techniques were considered to be more persuasive than those using reward-based techniques (loss-based emails: $M = 3.53$, $SE = .063$, $CI = 3.40, 3.65$; reward-based emails: $M = 3.87$, $SE = .047$, $CI = 3.78, 3.96$), supporting hypothesis 1a. An interaction between influence technique and authentic design cues was also found, $F(1, 176) = 5.154$, $p = .024$, $\eta 2 = .003$, such that the difference between loss- and reward-based emails was larger in the high authentic design cue condition compared to the low authentic design cue condition.

H1b: A main effect of influence technique was also found for ratings of email trustworthiness, $F(1, 176) = 11.49$, $p = .001$, $\eta 2 = .009$, such that emails using loss-based influence techniques were considered to be more trustworthy than those using reward-based techniques (loss-based emails: $M = 3.78$, $SE = .058$, $CI = 3.67, 3.90$; reward-based emails: $M = 3.98$, $SE = .043$, $CI = 3.89, 4.06$), supporting hypothesis 1b. Similarly to the results for email persuasiveness, an interaction was also found between influence technique and authentic design cues, $F(1, 176) = 12.873$, $p < .001$, $\eta 2 = .047$, such that the difference between loss- and reward-based emails was larger in the high authentic design cue condition compared to the low authentic design cue condition.

H1c: Finally, a main effect of influence technique was also found in relation to response choice, $F(1, 176) = 41.56$, $p < .001$, $\eta 2 = .108$, such that emails using loss-based influence techniques were more likely to be responded to than those using reward-based techniques (loss-based emails: $M = 1.63$ (63% of loss-based emails

rated ‘ignore’, 37% of loss-based emails rated ‘respond’), *SE* = .025, *CI* = 1.58, 1.68; reward-based emails: *M* = 1.83 (83% of reward-based emails rated ‘ignore’, 17% of reward-based emails rated ‘respond’), *SE* = .17, *CI* = 1.79, 1.86), supporting hypothesis 1c. An interaction between influence technique and authentic design cues was found, $F(1, 176) = 17.18$, $p < .001$, $\eta 2 = .026$, such that the difference between loss- and reward-based emails was larger in the high authentic design cues condition compared to the low authentic design cues condition.

3.2. H2: The Impact of Authentic Design Cues on Email Judgements

H2a: A significant main effect of authentic design cues was found for ratings of email trustworthiness, $F(1, 176) = 197.80$, $p < .001$, $\eta 2 = .307$, with emails in the high authentic design cue (HA) condition rated as more trustworthy than emails in the low authentic design cue (LA) condition (HA: *M* = 3.51, *SE* = .066, *CI* = 3.38, 3.64; LA: *M* = 4.48, *SE* = .041, *CI* = 4.40, 4.56), supporting hypothesis 2a, that emails containing authentic design cues would be rated as more trustworthy than emails that did not.

H2b: Similar to ratings of trust, a significant main effect of authentic design cues was also found for email persuasiveness, $F(1, 176) = 182.30$, $p < .001$, $\eta 2 = .406$, with emails in the high authentic design cue (HA) condition rated as more persuasive than emails in the low authentic design cue (LA) condition (HA: *M* = 3.46, *SE* = .062, *CI* = 3.34, 3.59; LA: *M* = 4.28, *SE* = .049, *CI* = 4.18, 4.37), supporting hypothesis 2b.

H2c: Finally, a significant main effect of authentic design cues was found for response choice, $F(1, 176) = 92.62$, $p < .001$, $\eta 2 = .173$, with emails in the high authentic design cues (HA) condition more likely to be responded to than emails in the low authentic design cues (LA) condition (HA: *M* = 1.70 (70% of HA emails rated

'ignore', 30% of HA emails rated 'respond'), *SE* = .024, *CI* = 1.66, 1.75; LA: *M* = 1.93 (93% of LA emails rated 'ignore', 7% of LA emails rated 'respond'), *SE* = .010, *CI* = 1.91, 1.95), supporting hypothesis 2c.

3.3. H3: The Impact of Salient Current Event Reference on Email Judgements

H3a: There was no significant effect of reference to salient current events on ratings of email trustworthiness $F(1, 176) = 1.17$, $p = .280$, $\eta 2 < .001$, therefore hypothesis 3a, that reference to salient current events would increase trust, was not supported. There was no significant interaction between salient current event reference and authentic design cues, $F(1, 176) = 3.38$, $p = .068$, $\eta 2 = .002$.

H3b: No significant effect of current event reference on ratings of email persuasiveness was shown $F(1, 176) = .013$, $p = .909$, $\eta 2 < .001$, so hypothesis 3b was also rejected. Again, there was no significant interaction between current event reference and authentic design cues, $F(1, 176) = 1.93$, $p = .17$, $\eta 2 = .001$.

H3c: Finally, no significant effect of salient current event reference was shown for response choice, $F(1, 176) = 2.79$, $p = .097$, $\eta 2 = .002$, so hypothesis 3c was also rejected. There was no interaction between salient current event reference and authentic design cues, $F(1, 176) = .31$, $p = .579$, $\eta 2 < .001$.

3.4. H4: The Impact of Online Risk Information on Email Judgements

No significant effect of group on either ratings of trust of emails (Risk information group *M* = 4.06, control group *M* = 3.93; *M Diff* = .130, *p* =.130, *SE* = .086, *CI* = -.039, .300) or ratings of persuasiveness (Risk information group *M* = 3.92, control group *M* = 3.82; *M Diff* = .09, *p* =.32, *SE* = .09, *CI* = -.091, .278) were found. Analysis of participant responses to two further self-report questions using independent-samples t-tests also showed that the groups did not differ on either of our two generic questions, namely 'To what extent do you trust email communications in

general?' (Risk information group $M = 3.29$, $SD = 1.57$, control group $M = 3.64$, $SD = 1.54$; $t(176) = 1.48$, $p = .14$) and 'How would you rate your awareness of the common techniques used in scams? (Risk information group $M = 4.75$, $SD = 1.77$, control group $M = 5.19$, $SD = 1.52$; $t(176) = 1.78$, $p = .08$). Therefore, hypothesis 4 was not supported.

Finally, to ensure that there was no difference across the two groups according to degree of familiarity with the four organisations used in the emails, which may have confounded any between group effects, a one-way ANOVA was conducted. This showed that mean ratings of familiarity for each of the four organisations within the emails did not differ across the two groups, all $F's(1, 176) < 2.42$, all $p's > .12$.

3.5. H5: The Relationship between Habitual Email Use and Trust

Responses on the Self-Report Habit Index ranged from 12.00 - 84.00 ($M = 42.75$, $SD = 16.88$). Self-reported habitual email use was found to be significantly related to an increased likelihood of choosing to respond to emails containing high authentic design cues across the various conditions (reference to Rio Olympics, $r = .163$, $p = .030$; no reference to Rio Olympics, $r = .171$, $p = .023$; loss-based, $r = .212$, $p = .005$) and low authentic design cue emails for two out of the three conditions (reference to Rio Olympics, $r = .169$, $p = .024$; no reference to Rio Olympics, $r = .002$, $p = .981$; loss-based, $r = .175$, $p = .020$). However, there was no significant relationship between habitual email use and ratings of email trust (all $r's < .13$, $p's > .1$), so hypothesis 5 was not supported.

3.6. What Factors did Participants Report as Influencing their Decision?

Table 3 presents a count of the most common categories identified within each of the three themes (i.e., influence techniques, authentic design cues, and habits and routines). This occurrence counting approach is comparable to the approach used by Williams et al. (2017) and not only aids in the identification of the key factors that participants consider to influence their decision, but also allows us to consider whether the predominant factors highlighted in self-reported data correspond to the actual results of the quantitative judgement data. The specific findings related to each theme are presented below. The response of one participant could reference multiple categories or themes.

Table 3. Three most common categories identified within each qualitative theme.

Themes					
Influence Techniques		**Authentic Design Cues**		**Habits and Routines**	
Suspicious of reward	27 times	Email address	59 times	Use alternative verification	9 times
Suspicious of personal information requests	21 times	Logo	51 times	Respond to only expected / usual emails	8 times
Professed organisation	14 times	Copyright statement	21 times	Avoid rewards / offers	7 times

3.6.1. The Role of Influence Techniques

Eighty-nine participants referenced the presence of particular influence techniques within the email as influencing their decisions. These influence techniques were considered to either encourage them to respond (48 participants), or conversely, to actually make them suspicious (41 participants). For instance, although the opportunity to access a potential benefit or reward was highlighted thirteen times as a reason to respond to an email *"the potential award or reimbursement for time"*, this

was also highlighted as something that raised suspicion twenty-seven times *"I would ignore… if it said you would win something or get something for free. Most emails like that are scams"*. In addition to the offer of a potential reward being considered suspicious, emails that contained a likely request for personal information within the professed scenario were also highlighted as less likely to be genuine twenty-one times *"if the email asked for personal information I wouldn't trust it"*. Other primary influence techniques that were considered by recipients to make them more likely to respond to an email were the reputation of the professed organisation that sent the email, *"what company sent the email"* (mentioned 14 times), and the perceived importance of responding to the email, *"if it involved identity then I would respond"* (mentioned 9 times).

3.6.2. The Presence of Authentic Design Cues

One hundred and four participants referenced authenticity cues as impacting their decision. These primarily related to the factors manipulated within the study, including the presence of a logo *"I feel like I'm more of a visual person, so whenever I see the symbol or something that indicates the email is trustworthy I answer it"* (51 times), the email address *"the determining factor (in trustworthiness) is the email address of the sender"* (59 times), and the presence of a copyright statement *"I think that the copyright at the end of the email […] made a difference in the way I would reply to the email"* (21 times). However, a number of responses also related to additional factors, such as the general layout of the email *"the format of the email"* (14 times), the wording used *"how the email was worded"* (11 times), and whether the email 'seemed legitimate' *"there is a 'look' to them, and when there is not something is off"* (8 times).

3.6.3. Individual Habits and Routine Behaviour

Thirty-three participants also discussed how they usually handled particular types of emails, reflecting aspects of their routine behaviour, habits and norms. For instance, always avoiding emails that offered some form of reward or offer *"I don't respond to emails that tell me to fill out a survey for a chance to win something. Usually because I never will win it anyway and also because it's a ploy to get your info out"* (7 times), those that included survey links *"I am the type of person that does not take online surveys that come through my email"* (5 times), or those that requested personal information *"I would hesitate to respond to anything asking for updated information, or personal information in general"* (6 times). This also included participants who purported always using alternative verification strategies when such emails were received *"I typically would see stuff like that and not click the link directly but rather go to the site itself to see if I could find it if I questioned its validity"* (9 times), and only ever responding to certain types of emails that were expected or usual *"I do not generally respond to any emails that do not pertain directly to school or work"* (8 times).

4. Discussion

The current study examined the potential factors that influence how individuals evaluate suspicious emails, in particular the extent to which they trust such emails, how persuasive they find them, and whether they would choose to respond to them. Specifically, a range of factors were examined that have previously been highlighted in theoretical models of phishing susceptibility (Vishwanath et al., 2011; Vishwanath et al., 2016), focusing on (a) aspects of the message itself, including whether it used *loss or reward-based* influence techniques, whether it contained particular *authentic design* cues, and whether it referenced a salient *current event* (the Rio Olympics), and (b) aspects outside of the message, in particular whether participants had been

exposed to prior news information highlighting the risks of online scams and the extent to which recipients engaged in habitual email use.

4.1. Message-Specific Factors

4.1.1. Loss and Reward-based Influence Techniques

Within the current study, emails that used loss-based influence techniques were found to be considered more trustworthy and more persuasive than emails focused on reward-based techniques, suggesting that the concept of loss aversion, whereby people are more sensitive to losses than gains (Kahneman & Tversky, 1984), may be applied to differential responding to phishing emails that use these techniques. People have also been suggested to be more willing to take a risk to avoid a potential loss than to achieve a gain (Kahneman & Tversky, 1979), which may result in recipients of suspicious emails being more likely to undertake risky actions, such as clicking on links, for loss-based emails.

In this way, emails that reference a potential loss to the individual if they fail to respond are considered to be more important to respond to, and also, interestingly in our study, more trustworthy, than those that offer a potential reward for responding. As shown in our findings for perceived persuasiveness, people may feel that loss-based emails are more persuasive and therefore more important to respond to and as a result, strive to remain consistent in their judgements by rating emails that they feel more pressured (and likely) to respond to as also being more trustworthy. Such processes can be attributed to individuals' motivation to avoid cognitive dissonance (Festinger, 1962), whereby people attempt to avoid contradictions within their thoughts and beliefs.

Alternatively, findings of greater relative trust for loss-based emails could simply be a result of greater suspicion towards reward-based emails. This may be due

to increasing publicity and awareness of scams that use reward-based techniques (such as the so-called 'Nigerian Prince' scam, where large amounts of money are offered) resulting in greater suspicion of communications that appear to offer 'something for nothing'. However, this appears to contrast with the findings of Harrison et al. (2016), whereby the presence of threat and reward-based cues did not differentially impact message processing or phishing susceptibility. It is possible therefore that these influence techniques may impact people's evaluations and perceptions more than their actual response behaviour and disentangling these processes would be a fruitful avenue for future research. Understanding this mechanism further is crucial if potential mitigations against this effect are to be developed, in particular whether interventions should focus on reducing public assumptions of trust in such emails or should instead focus their efforts on reducing the persuasive effect of this influence technique.

4.1.2. Authentic Design Cues

Overall, results showed that participants were more likely to consider phishing emails that included authentic design cues, such as logos and copyright statements, as more trustworthy and more persuasive than those that did not. This suggests that individuals base their evaluations of the legitimacy of emails that they receive on the presence of such cues, as proposed by Sillence et al. (2006) in their consideration of judgements of health-related websites. This may occur because emails that contain logos, copyright statements, and include a familiar brand name within the sender address, invoke particular stereotypes and expectations regarding what a 'genuine' email 'looks like', resulting in such emails failing to invoke contextual suspicion (where suspicion is evoked by the stimuli itself). Without a reason to doubt the legitimacy of an email, participants may then simply defer to assuming that the communication is

likely to be genuine. This is in line with literature regarding the existence of a truth bias in the general population, where people default to assuming that information is likely to be true unless they have a reason to doubt it (Levine, 2014).

Similarly, increased trust in emails containing authentic design cues could also be considered in relation to the concept of integrity-based trust. Integrity relates to whether one adheres to a set of acceptable principles (Mayer, Davis & Schoorman, 1995), and previous research suggests that violations of integrity can substantially reduce perceived trustworthiness (Kim, Dirks, Cooper & Ferrin, 2006). These same principles may thus apply to email communications, such that authentic design cues represent expected and acceptable principles of legitimate communications, with violations related to these elements (e.g., lacking a copyright statement or a particular structure of sender address) leading to decreased trust in those communications. Alternatively, the presence of such cues may monopolise attentional resources when participants are considering how trustworthy the email is, in a similar way to that proposed by Vishwanath et al. (2011) regarding the capability of influence techniques to monopolise attention when people decide to respond to a phishing email. In the former scenario, the absence of authentic design cues may actively violate integrity expectations, leading to a decrease in relative trust compared to a 'default trust' baseline. In the latter, the presence of authentic design cues instead attract attention and lead to increased ratings of trust relative to a more suspicious evaluative baseline. The primary difference between these possibilities being (a) whether a default trust baseline is assumed and then violated, resulting in decreased trust or (b) whether the presence of these cues increases trust levels relative to a lower baseline level. Further work to understand these potential mechanisms using, for instance, eye tracking approaches and reaction time studies, would be beneficial.

Unfortunately, from a practical perspective, each of these design elements are currently used within phishing emails to persuade people to respond. As a result, it is vital that effective means of assisting individuals to circumvent potentially automatic responses and assumptions related to such stimuli are developed. This presents a substantial challenge that will likely require a combination of education (Abawajay, 2014), training (Sheng, Magnien, Kumaraguru, Acquisti, et al., 2007) and interface design approaches (Mohamed, Chakraborty, & Dehlinger, 2016).

4.1.3. Referencing Salient Current Events

A number of scam communications have emerged that reference current events, however to our knowledge, no research has examined whether linking phishing emails to events that individuals are familiar with influences trust in, or persuasiveness of, those emails. This is despite previous research suggesting that previously encountered information is more likely to be considered genuine (Begg et al., 1992). Within the current study, we primed participants with a current event (the 2016 Rio Olympics) and referenced this event within a subset of emails to see whether referencing salient current events increased the perceived trustworthiness and persuasiveness of emails. However, no significant effect of trust, persuasiveness or response likelihood was found, suggesting that prior exposure to information regarding current events does not make phishing emails that link to such events appear any more legitimate than those that do not.

It is possible that this lack of effect was due to a general suspicion of all reward-based emails, with the presence of such influence techniques effectively overriding any potential effect of salient current event reference. Indeed, participants rated reward-based emails as less trustworthy overall than those using loss-based influence techniques. Alternatively, participants may not have been sufficiently

interested in the offers provided across any of the reward-based emails, leading to a masking of potential effects related to current event reference. This may account for the lower ratings of persuasiveness of reward-based emails more generally. However, this explanation does not account for lower ratings of trustworthiness in such emails. In order to ensure that the number of emails participants evaluated was not too high, we did not include any reference to salient current events within the loss-based emails. However, to investigate the likely mechanisms related to our findings further, particularly whether any current event reference effects may have been masked by the use of reward-based influence techniques, it would be useful for future research to explore the impact of salient current event references on judgements of loss-based emails (e.g., using loss-based emails that reference recent high-profile data breaches in relation to account suspensions etc.).

Finally, it is possible that the current event reference lost its salience over the course of the email judgement task. However, since the task used was relatively short (approximately 15-20 minutes duration) and individuals are unlikely to receive a phishing email at precisely the same time, or within a few minutes, of being exposed to relevant current event information, if this was the case then this would suggest that referencing current events would not realistically impact susceptibility within real world contexts.

4.2. External Factors

4.2.1. Exposure to Generic Risk Information

The impact of exposure to media information regarding the risks of online scams was also examined within the current study. It was hypothesised that individuals who were exposed to generic information that highlighted the risks of online communications would show decreased ratings of trust and persuasiveness due to heightened

perceptions of the phishing risk and increased generalised suspicion. However, exposure to such information did not significantly impact either trust ratings or persuasiveness. This suggests that exposure to media reports or other information that aims to raise general awareness of online threats may not be sufficient to impact resultant susceptibility.

Indeed, we purposefully did not include any particular information regarding what makes emails suspicious, or particular cues to look out for, as has been the case in previous education intervention approaches (e.g., Abawajay, 2014). This allowed us to investigate whether generic information that related to potential online risks, but did not include any specific advice, would be sufficient to make people more suspicious of communications that contained suspicious elements. Since individuals are increasingly bombarded with protective information relevant to all aspects of their lives, further work that systematically manipulates the amount and type of intervention information that people are exposed to would be beneficial. Within this study, the intervention manipulation used was not sufficient to influence judgements and inclusion of additional information or advice may be necessary to show any form of effect. Further manipulation of the depth and type of information presented in the future would allow greater understanding of how much information is sufficient, and how this information is best presented, to maximise the likelihood that people will identify fraudulent communications.

4.2.2. Email Habits

In addition to manipulating specific aspects of the phishing email that is viewed, this study also explored the impact of factors related to the individual who views the message. Overall, habitual email use was significantly related to an increased likelihood of choosing to respond to emails than to ignore them. This supports

previous findings by Vishwanath (2015) that email habits can increase participant susceptibility to a simulated phishing attack. However, this study also extends these findings by demonstrating that (a) this relationship occurs across different types of phishing emails, and (b) the increased likelihood of response does not correspond to increased ratings of trust in such emails. Therefore, although people who engage in habitual email use appear to be more likely to choose to respond to phishing emails, they do not consider these emails to be any more trustworthy than individuals who do not engage in habitual email use. This suggests that the increased likelihood of responding to emails in such individuals is due to automatic response processes rather than biased trust assessments arising from habitual email use.

4.3. Limitations

There are a number of limitations within the current study that may limit the conclusions that can be made from the data. Firstly, the participant sample consisted of university students within a relatively small age range. This does not represent the full range of email users and it is possible that individuals with other demographic characteristics may evaluate our message-specific factors in different ways. For instance, younger adults (18-25 years) have been identified as more susceptible to phishing websites than older adults. However, this effect is considered to be moderated by educational background, with greater education reducing susceptibility (Gavett, Zhao, John, Bussell, et al, 2017; Sheng, Holbrook, Kumaraguru, Cranor, et al, 2010). Although the reasons for this are uncertain, factors associated with greater critical thinking skills and learning and cognition have been highlighted. Therefore, it is possible that other population samples will be more, or differentially, susceptible to the influence techniques used within the current study. Further work exploring the extent to which email judgements are similar, or indeed differ, across different age

ranges and educational backgrounds, as well as other demographic characteristics, would therefore be beneficial.

Secondly, in order to compare a number of email elements, it was necessary for participants to view, and make judgements of, a number of emails. This prevented the use of a phishing simulation approach, since it would not have been practical to send 20 phishing emails to each participant. As a result, actual response behaviour could not be measured across the different email types, instead participant self-report perceptions were relied upon. However, this did allow for the collection (and therefore separation) of data related to both degree of trust and persuasiveness and likelihood of response, which enabled a separate consideration of these potential mechanisms. After all, increased trust in an email does not necessarily mean that recipients will be persuaded to respond. Similarly, participants may be motivated to respond to a particular email despite potential misgivings about its authenticity.

It is also possible that asking participants the extent to which they trusted each email may have itself invoked a degree of suspicion regarding potential message legitimacy. It should, therefore, be considered that this could have reduced the impact of the online risks news story aspect of this study, due to suspicion already being invoked in both groups. Further work could examine this possibility. However, if this is the case, then it would suggest that general awareness material does not impact response behaviour to any greater degree than simply requiring people to indicate the degree to which they think a message is trustworthy at the time that the message is processed. If so, this could itself represent a future mitigation strategy.

Finally, the email stimuli that were used within the current study did not allow participants to interact with the email in any way (i.e., they were unable to hover over the link to verify the URL that it led to). However, this approach did enable us to

explore how the other factors under study influenced participant perceptions, providing the foundation for further work to build upon these aspects and identify the potential contributions of more interactive strategies and email elements in the future.

5. Conclusions

Recent advances in phishing susceptibility research have expanded current understanding of how people make decisions regarding suspicious emails. However, the precise role of various message-specific factors, including how and why they influence people's judgements and decisions, remains unclear. The current study investigated how three of these factors, which have not been extensively examined in previous research, influence judgements of email trust and persuasiveness, specifically the use of loss and reward-based influence techniques, authentic design cues, and referencing a salient current event. The use of loss-based influence techniques and the presence of authentic design cues was found to increase perceived trust and persuasiveness, with a number of psychological mechanisms identified that may account for these findings. It is hoped that these findings will provide a basis from which to systematically explore the potential role of these various underlying mechanisms in the future. Only by understanding *how* people evaluate email communications more generally will it be possible to understand precisely *why* phishing emails work and how best to mitigate them.

References

Atkins, B., & Huang, W. (2013). A study of social engineering in online frauds. *Open Journal of Social Sciences, 1, 03,* 23-23. DOI: 10.4236/jss.2013.13004.

Abawajy, J. (2014). User preference of cyber security awareness delivery methods. *Behaviour & Information Security, 33*(3)*,* 237-248. DOI: 10.1080/0144929X.2012.708787.

Arachchilage, N.A.G., & Love, S. (2014). Security awareness of computer users: A phishing threat avoidance perspective. *Computers in Human Behaviour, 38,* 304-312. DOI: 10.1016/j.chb.2014.05.046.

BBC News. (2017). *UK tourists hit by booking scams up by nearly a fifth*. Cited from http://www.bbc.co.uk/news/uk-39835874 on 7th May 2017.

BBC News. (2016). *Students warned of new 'phishing' scam.* Cited from http://www.bbc.co.uk/news/education-37408373 on 30th January 2017.

Begg, I.M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*(4)*,* 446-458. DOI: 10.1037/0096-3445.121.4.446.

Boureau, Y.L., Sokol-Hessner, P., & Daw, N.D. (2015). Deciding how to decide: self-control and meta-decision making*. Trends in Cognitive Science, 19*(11)*,* 700-710. DOI: 10.1016/j.tics.2015.08.013.

Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A. (2015). Breaching the human firewall: Social engineering in phishing and spear phishing emails. *Australasian Conference on Information Systems.* arXiv:1606.00887.

Canfield, C., Fischhoff, B., & Davis, A. (2016). Quantifying phishing susceptibility for detection and behavior decisions. *Human Factors, 58*(8)*,* 1158-1172. DOI: 10.1177/0018720816665025.

Cialdini, R. (2007). *Influence: The psychology of persuasion*. New York: HarperCollins.

Festinger, L. (1962). Cognitive dissonance, *Scientific American*, *207*(4), 93–107. DOI: 10.1038/scientificamerican1062-93.

Fogg, B.J., Marshall, J., Ospiovich, A., Varma, C., Laraki, O., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2000). Elements that affect web credibility: early results from a self-report study. *Proceedings of CHI '00 Extended Abstracts on Human Factors in Computing Systems,* (pp. 287-288). DOI: 10.1145/633292.633460.

Gavett, B.E., Zhao, R., John, S.E., Bussell, C.A., Roberts, J.R., & Yue, C. (2017). Phishing suspiciousness in older and younger adults: The role of executive functioning. *PLOS One.* DOI: 10.1371/journal.pone.0171620.

Harrison, B., Svetieva, E., & Vishwanath, A. (2016). Individual processing of phishing emails: How attention and elaboration protect against phishing. *Online Information Review, 40*(2), 265-281. DOI: 0.1108/OIR-04-2015-0106.

Harrison, B., Vishwanath, A., & Rao, R. (2016). A user-centered approach to phishing susceptibility: the role of a suspicious personality in protecting against phishing. In *49th Hawaii International Conference on System Sciences*, 5th-8th Jan. DOI: 10.1109/HICSS.2016.696.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica. 47*(2), 263. DOI: 10.2307/1914185.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341-350. DOI: 10.1037/0003-066X.39.4.341.

Kim, P.H., Dirks, K.T., Cooper, C.D., & Ferrin, D.L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence-vs. integrity-based trust violation. *Organizational Behaviour and Human Decision Processes, 99*(1), 49-65.

Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L.F., & Hong, J. (2007). Getting users to pay attention to anti-phishing education: evaluation of retention and transfer. *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*, 4th-5th Oct, 70-81. DOI: 10.1145/1299015.1299022.

Lang, A. (2000). The Limited Capacity Model of Mediated Message Processing. *Journal of Communication, 50*(1)*,* 46-70. DOI: 10.1111/j.1460-2466.2000.tb02833.x.

Levine, T. R. (2014). Truth-default theory: A theory of human deception and deception detection. *Journal of Language and Social Psychology, 33,* 378-392. DOI: 10.1177/0261927X14535916.

Mayer, R.C., Davis, J.H., & Schoorman, F.D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709-734.

Metzger, M.J., Flanagin, A.J., & Medders, R.B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication, 60*(3), 413-439. DOI: 10.1111/j.1460-2466.2010.01488.x.

Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90,* 227–234. DOI: 10.1037/h0031564.

Modic, D., & Anderson, R. J. (2014). Reading this may harm your computer: The psychology of malware warnings. *Computers in Human Behavior, 41,* 71-79. DOI: 10.1016/j.chb.2014.09.014.

Mohamed, M.A., Chakraborty, J., & Dehlinger, J. (2016). Trading off usability and security in user interface design through mental models. *Behaviour and Information Technology,* 1-24. DOI: 10.1080/0144929X.2016.1262897.

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2013). Phishing for the truth: a scenario-based experiment of users' behavioural response to emails. In Janczewski L.J., Wolfe H.B., Shenoi S. (eds.), *Security and Privacy Protection in Information Processing Systems. SEC 2013. IFIP Advances in Information and Communication Technology,* vol 405. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-39218-4_27.

Pattinson, M., Jerram, C., Parsons, K.M., McCormac, A., & Butavicius, M.A. (2012). Why do some people manage phishing emails better than others? *Information Management & Computer Security, 20*(1), 18-28. DOI: 10.1108/09685221211219173.

Oliveira, D., Rocha, H., Yang, H., Ellis, D., Dommaraju, S., Muradoglu, M., Weir, D., Soliman, D., Lin, T., Ebner, N. (2017). Dissecting spear phishing emails for older vs young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17,* 6412-6424. DOI: 10.1145/3025453.3025831.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L.F., Downs, J. (2010). Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions. *Proceedings of the 2010 CHI Conference on*

Human Factors in Computing Systems, CHI ' 10, 373-382. DOI: 10.1145/1753326.1753383.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J., & Nunge, E. (2007). Anti-phishing Phil: the design and evaluation of a game that teaches people not to fall for phish. *Proceedings of the 3rd symposium on Usable privacy and security (SOUPS),* 88-99. DOI: 10.1145/1280680.1280692.

Sillence, E., Briggs, P., Harris, P., & Fishwick, L. (2006). A framework for understanding trust factors in web based health advice. *International Journal of Human Computer Studies, 64,* 697-713. DOI: 10.1016/j.ijhcs.2006.02.007.

Stajano, F., & Wilson, P. (2011). Understanding scam victims: Seven principles for systems security. *Communications of the ACM*, *54(3),* 70-75. DOI: 10.1145/1897852.1897872.

Sun, J.C-Y., & Chen, A.Y-Z. (2016). Effects of integrating dynamic concept maps with Interactive Response System on elementary school students' motivation and learning outcome: The case of anti-phishing education. *Computers & Education, 102,* 117-127. DOI: 10.1016/j.compedu.2016.08.002.

Toma, C.L., & Hancock, J.T. (2012). What lies beneath: the linguistic traces of deception in online dating profiles. *Journal of Communication, 62,* 78-97. DOI: 10.1111/j.1460-2466.2011.01619.x.

Verizon. (2016). 2016 Data breach investigations report. Accessed on 23.09.2016 at http://www.verizonenterprise.com/verizon-insights-lab/dbir/.

Verplanken, B., & Orbell, S. (2003). Reflections on past behaviour: A self-report index of habit strength. *Journal of Applied Social Psychology, 33*(6), 1313-1330.

Vishwanath, A. (2015). Habitual Facebook use and its impact on getting deceived on social media. *Journal of Computer-Mediated Communication, 20,* 83-98. DOI: 10.1111/jcc4.12100.

Vishwanath, A., Harrison, B., & Ng, Y. J. (2016). Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication Research, online pre-print,* 1-21. DOI: 10.1177/0093650215627483.

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H. R. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems, 51,* 576-586. DOI: 10.1016/j.dss.2011.03.002.

Williams, E.J., Morgan, P., & Joinson, A. (2017). Press accept to update now: Individual differences in susceptibility to malevolent computer updates. *Decision Support Systems, 96,* 119-129. DOI: 10.1016/j.dss.2017.02.014.

Workman, M. (2008). Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security. *Journal of the American Society for Information Science and Technology, 59*(4), 662–674. DOI: 10.1002/asi.20779.

Wright, R.T., & Marett, K. (2010). The influence of experiential and dispositional factors in phishing: An empirical investigation of the deceived. *Journal of Management Information Systems, 27*(1), 273-303. DOI: 10.2753/MIS0742-1222270111.

Appendix A. News stories provided to participants in risk information group and control group prior to email judgement task.

Rio Olympics 2016: Spectacular closing ceremony *(both groups)*

The spectacular ending of this summer's Olympic Games in Rio de Janeiro featured a colorful carnival parade and ceremony lasting almost 3 hours. The 16-day games, which showcased over 11,000 athletes, were officially closed by the chief of the International Olympic Committee, Thomas Bach. The closing ceremony on Aug. 21. was watched by billions of people around the world and included a dramatic extinguishing of the Olympic flame.

The rise in online scams *(risk information group only)*

A number of public-safety forums have been established to warn residents of the dangers of scams in the digital age. Criminals are making use of advancements in technology to make scams increasingly targeted and convincing. Anyone can fall victim to a scam, young or old. Across the world the problem of fake websites, emails and fraudulent phone calls is on the rise, so people need to increasingly be on their guard in this technological era.

The challenge of sedentary lifestyles *(control group only)*

Sitting down all day is not good for you. The rise in desk jobs and inactivity means that many of us have a sedentary lifestyle. Research has shown that the long periods of physical inactivity associated with a sedentary lifestyle are bad for our health. Reducing this risk remains a challenge across the globe, but one thing is for sure, we should all be trying to keep ourselves moving -so stand up, move more and 'watch' less.

Appendix B. Example stimuli for each of the email conditions.

a) High authenticity email related to Rio Olympics.

From: [Company Name] (do_not_reply@rewards.[company name].com)
Date: Wed, 22 Jun 2016 12:23:55 +0200
Subject: Win free flights to celebrate Rio 2016!

Company Logo **Rio Logo**

Customer service at [company name] is focused on spirit and warmth. As proud supporters of the spirit of the **Rio Olympics**, [company name] are offering a limited number of free flights across 2017 so that you can share the spirit too!

Just click HERE to be in with a chance to win.

Thank you for supporting [company name]!

Sincerely,

[Company name] Rewards

b) High authenticity email unrelated to Rio Olympics.

From: [CompanyName](do_not_reply@rewards.[companyname].com)
Date: Wed, 22 Jun 2016 12:23:55 +0200
Subject: You could win a $500 Gift card: Fill out our survey!

Company Logo

At [company name], we are dedicated to the best-quality Customer Service delivered with a sense of friendliness, company spirit and warmth. Give us your thoughts now to be in with a chance to win a $500 voucher to use on your next flight!

Just click HERE to access our survey today.

Thank you for supporting [company name]!

Sincerely,

[company name] Rewards

c) Low authenticity email related to Rio Olympics.

From: customersurvey@gmail.com
Date: Wed, 22 Jun 2016 12:23:55 +0200
Subject: Win free flights to celebrate Rio 2016!

[Company name] have always focused customer service on spirit and warmth. As proud supporters of the spirit of the **Rio Olympics**, we are offering a limited number of free flights across 2017 so that you can share the spirit too!

Just click HERE to be in with a chance to win.

Thank you for supporting [company name]!

Sincerely,

[Company name] Rewards

d) Low authenticity email unrelated to Rio Olympics.

From: customersurvey@gmail.com
Date: Wed, 22 Jun 2016 12:23:55 +0200
Subject: You could win a $500 Gift card: Fill out our survey!

[Company name] are dedicated to providing the best-quality Customer Service delivered with a sense of friendliness, company spirit and warmth. Just give us your thoughts to be in with a chance to win a $500 voucher to use on your next flight!

Just click HERE to access our survey today.

Thank you for supporting [company name]!

Sincerely,

[Company name] Rewards

e) High authenticity loss email.

From: [Company Name] (do_not_reply@live.[company name].com)
Date: Wed, 22 Jun 2016 12:23:55 +0200
Subject: Important Information: Account suspension

Company Logo

Hello,

Your internet banking access has been temporarily suspended in order to protect your account against possible misuse. **In order to reinstate your [company name] account, please verify your details at the link below as soon as possible.**

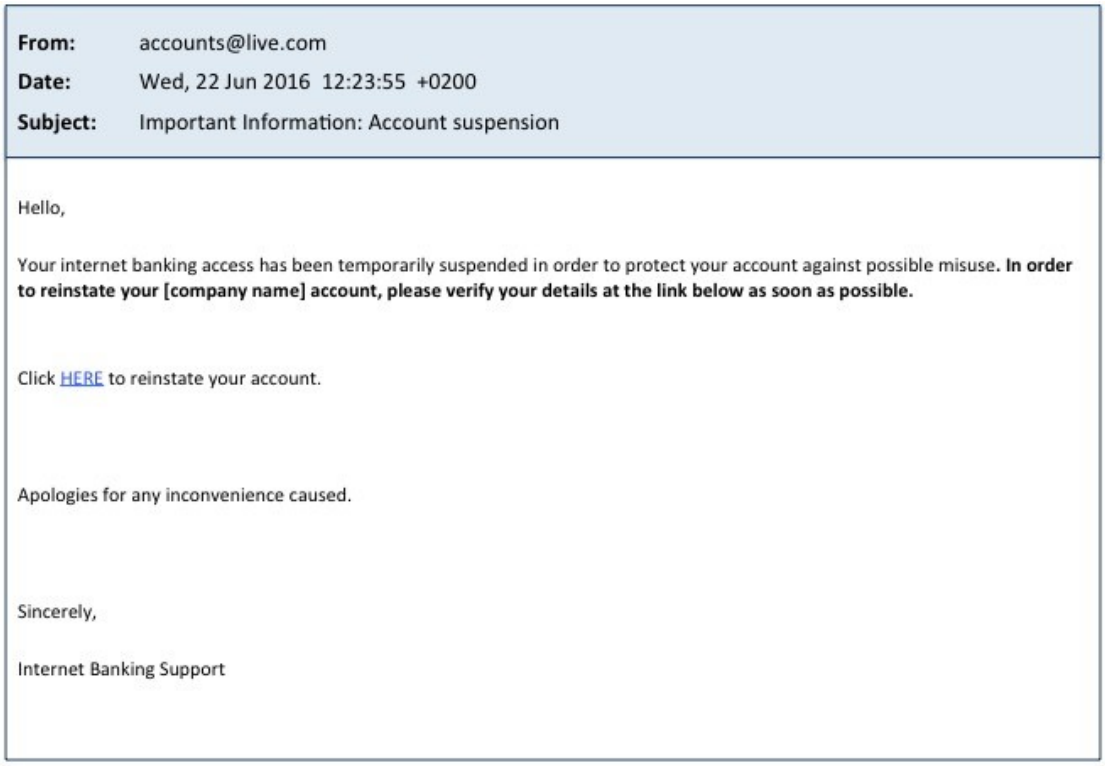
Click HERE to reinstate your account.

Apologies for any inconvenience caused.

Sincerely,

[company name] Internet Support

f) Low authenticity loss email.



From: accounts@live.com
Date: Wed, 22 Jun 2016 12:23:55 +0200
Subject: Important Information: Account suspension

Hello,

Your internet banking access has been temporarily suspended in order to protect your account against possible misuse. **In order to reinstate your [company name] account, please verify your details at the link below as soon as possible.**

Click HERE to reinstate your account.

Apologies for any inconvenience caused.

Sincerely,

Internet Banking Support